

# PENERAPAN DATA MINING MENGGUNAKAN METODE TEKNIK CLASSIFICATION UNTUK MELIHAT POTENSI KEPATUHAN WAJIB PAJAK BUMI DAN BANGUNAN

**Qoriani Widayati**  
**Dosen Universitas Bina Darma**  
**Jalan Jenderal Ahmad Yani No.3 Palembang**  
**Sur-el: qoriani\_widayati@binadarma.ac.id**

---

**Abstract:** *The government implements development in Indonesia, requires substantial funds. The entry of cash from the Land and Building Tax is the most important part for the development of a region, with the results that have been obtained by the regional government can increase regional development with various infrastructures that help the community to carry out various activities and make the area more advanced. One type of tax is the Land and Building Tax (PBB). With the increasing number of taxpayers and data paying contributions directly into the treasury of state finances, the UPT BPPD of SU II Subdistrict of Palembang city did not know how many obedient and non-compliant taxpayers. In this study using data mining techniques, namely classification by applying the Naive Bayes algorithm and getting from the number of taxpayers as many as 1,647 taxpayers with an accuracy of 99.33% which has the potential to not be on time in 16 ulu villages at 0,437 and sub-district households with data of 0.229.*

**Keywords:** *United Nations, Data Mining, taxpayers*

**Abstrak:** *Pemerintah melaksanakan pembangunan di Indonesia ini tentu memerlukan dana yang cukup besar. Pemasukan kas dari Pajak Bumi dan/atau Bangunan merupakan bagian terpenting untuk pembangunan suatu daerah, dengan hasil yang telah didapatkan pemerintah daerah dapat meningkatkan pembangunan daerah dengan berbagai infrastruktur yang membantu masyarakat dalam melakukan berbagai aktivitas serta menjadikan daerah tersebut menjadi lebih maju. Salah satu jenis pajak adalah Pajak Bumi dan Bangunan (PBB). Dengan semakin banyaknya wajib pajak dan data yang membayar iuran langsung masuk ke kas keuangan negara menyebabkan pihak UPT BPPD Kecamatan SU II kota Palembang tidak mengetahui seberapa banyak wajib pajak yang patuh dan tidak patuh. Pada penelitian ini menggunakan teknik data mining yaitu classification dengan menerapkan algoritma naive bayes dan mendapatkan dari jumlah wajib pajak sebanyak 1.647 wajib pajak dengan akurasi sebesar 99.33% yang berpotensi tidak tepat waktu terdapat pada kelurahan 16 ulu sebesar 0.437 dan kelurahan tangga takat dengan data sebesar 0.229*

**Kata kunci:** *PBB, Data Mining, wajib pajak*

---

## 1. PENDAHULUAN

Pemerintah melaksanakan pembangunan di Indonesia ini tentu memerlukan dana yang cukup besar. Pemasukan kas dari Pajak Bumi dan/atau Bangunan merupakan bagian terpenting untuk pembangunan suatu daerah, dengan hasil yang telah didapatkan pemerintah daerah dapat meningkatkan pembangunan daerah dengan berbagai infrastruktur yang membantu masyarakat dalam melakukan berbagai aktivitas

serta menjadikan daerah tersebut menjadi lebih maju. Pajak Bumi dan Bangunan (PBB) adalah pajak negara yang dikenakan terhadap bumi dan bangunan berdasarkan undang-undang nomor 12 tahun 1985 tentang pajak bumi dan bangunan (PBB) adalah pajak yang bersifat kebendaan dalam arti besarnya pajak terutang ditentukan oleh keadaan objek yaitu bumi/tanah dan atau bangunan. Keadaan subjek (siapa yang membayar) tidak ikut menentukan besarnya pajak ([www.pajak.go.id](http://www.pajak.go.id)).

Kepatuhan yang tinggi dari wajib pajak adalah suatu keadaan dimana wajib pajak memenuhi semua kewajibannya dalam membayar pajak, namun dalam realitanya tidak semua bahkan rendahnya tingkat wajib pajak patuh dalam membayar pajak.

Dinas Pendapatan Daerah adalah suatu instansi yang bertujuan untuk mengumpulkan penerimaan Pendapatan Asli Daerah (PAD) yang bersumber dari pajak daerah lainnya untuk pembangunan daerah. Terhitung mulai tanggal 01 Januari 2017 Dinas Pendapatan Daerah Kota Palembang berubah menjadi Badan Pengelolaan Pajak Daerah (BPPD) Kota Palembang berdasarkan Peraturan Pemerintah Nomor 18 Tahun 2016.

Dengan melihat kemajuan infrastruktur kota Palembang yang semakin maju menjadikan pajak bumi dan bangunan (PBB) sebagai pendapatan asli daerah (PAD) yang paling dominan, namun yang menjadi masalahnya dengan menjadikan PBB sebagai salah satu penyumbang terbesar PAD yang menyebabkan peningkatan jumlah wajib pajak bumi dan bangunan semakin terus mengalami kenaikan, sehingga secara otomatis membuat masyarakat menjadi hilang kesadaran dan membuat menurunnya kepatuhan wajib pajak dalam membayar iuran tersebut. Dengan semakin meningkatnya jumlah wajib pajak dan sistem pembayaran melalui bank dengan iuran yang langsung masuk ke data kas keuangan negara menyebabkan pihak UPT BPPD Kecamatan SU II tidak mengetahui seberapa banyak wajib pajaknya yang patuh dan seberapa banyak wajib pajak yang tidak patuh dalam membayar iuran pajak bumi dan bangunan (PBB) tersebut. Dari

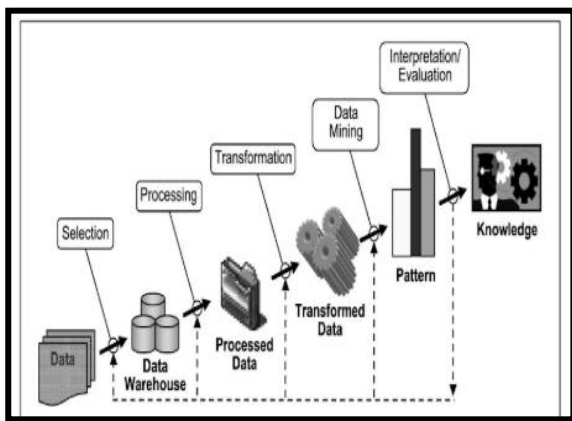
hal tersebut dilihat manfaat data mining adalah untuk menangani banyaknya data. Untuk melihat kepatuhan WP digunakan teknik klasifikasi untuk menentukan potensi kepatuhan wajib pajak (WP). Jumlah wajib pajak untuk tahun 2017 sebanyak 13.977 WP pada Kecamatan Seberang Ulu 2 Palembang. Data Mining adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat pada basis data. Data mining terutama digunakan untuk mencari pengetahuan yang terdapat dalam basis data yang besar sehingga sering disebut *Knowledge Discovery Database (KDD)* (Vulandari, 2017).

## 2. METODOLOGI PENELITIAN

Metode yang digunakan dalam penelitian ini adalah metode deskriptif. Metode deskriptif adalah suatu metode dalam meneliti suatu objek, suatu metode yang mengemukakan masalah dengan mengumpulkan data-data yang bertujuan untuk mendapatkan gambaran yang jelas mengenai suatu keadaan dengan cara menyajikan. penelitian deskriptif yaitu, penelitian yang dilakukan untuk mengetahui nilai variabel mandiri, baik satu variabel atau lebih (independen) tanpa membuat perbandingan, atau menghubungkan dengan variabel yang lain (Sugiyono, 2012).

Pada Pada metode analisis akan dimining dengan melalui tahapan *knowledge discovery in database* (KDD). Proses KDD adalah proses menggunakan data mining untuk mengekstrak pengetahuan apa yang dianggap sesuai dengan spesifikasi ukuran dan batas, menggunakan *database* bersama dengan *preprocessing* yang diperlukan, pengambilan sampel dan transformasi dari *database*.

Adapun untuk menganalisis data dalam penerapan data mining ini menggunakan tahapan *Knowledge Discover in Database* (KDD). istilah data mining dan *knowledge discovery in database* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi yang tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan dengan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining (Nofriansyah, 2014).



**Gambar 1. Proses KDD**

Berdasarkan gambar di atas, proses KDD secara garis besar dijelaskan sebagai berikut:

a. *Data Selection*

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, kata memilih data-data seperti apa saja yang kita butuhkan untuk proses lebih lanjut dan kemudian data disimpan dalam suatu berkas, terpisah dari basis data operasional sehingga memberikan kemudahan untuk penggunaan berikutnya.

b. *Pre-processing (Cleaning)*

Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut data yang tidak relevan itu juga lebih baik dibuang karena keberadaannya bisa mengurangi mutu atau akurasi dari hasil data mining nantinya.

c. *Data Transformation*

Beberapa teknik data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa teknik standar seperti analisis asosiasi dan klustering hanya bisa menerima *input* kategorikal. Karenanya data berupa angka numerik yang berlanjut bisa dibagi-bagi menjadi beberapa interval. Proses ini sering disebut *binning*. Disini juga dilakukan pemilihan data yang diperlukan oleh teknik data mining yang dipakai. Transformasi dan pemilihan data ini juga menemukan kualitas dari hasil data mining nantinya karena ada beberapa karakteristik dari teknik-teknik data mining tertentu tergantung pada tahap ini.

d. *Data Mining*

Data mining adalah proses menjadi pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat tergantung pada tujuan dan proses KDD secara keseluruhan. Data yang digunakan untuk bisa menjadi sebuah model yang baik idealnya mencukupi sebagai data riset. Semakin banyak data dan semakin sedikit kesalahan (*error*) maka semakin bagus model yang dijadikan patokan.

e. *Interpretation (Evaluation)*

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta pola atau informasi yang ada sebelumnya.

Data yang didapatkan dari UPT BPPD kota Palembang yaitu data PBB tahun 2017 pada kecamatan seberang ulu 2 Palembang dengan keseluruhan data yang telah membayar iuran PBB sebanyak 5.284 WP.

Sebelum dilakukannya proses KDD data akan dihapuskan jumlah tagihan dibawah Rp. 100.000 sesuai dengan peraturan pemerintah kota Palembang tahun 2018 dengan ditiadakannya iuran untuk biaya dibawah Rp. 100.000. Proses penghapusan data tersebut akan dilakukan menggunakan *software pentaho* yang akan dijelaskan selanjutnya, setelah data tidak

mengandung tagihan dibawah Rp. 100.000 data tersebut akan diproses pada tahapan KDD dengan tahapan yaitu seleksi data, *pre-processing* data, transformasi data, data mining dan evaluasi.

### 3. HASIL DAN PEMBAHASAN

Adapun untuk menganalisis data dalam penerapan data mining ini menggunakan tahapan *Knowledge Discover in Database (KDD)* yang sebelumnya akan dipisahkan menggunakan *tools Pentaho*, dengan tahapan sebagai berikut:

- 1) *Remove Data*  $\leq$  Rp. 100.000 menggunakan *Pentaho*
- 2) *Data Selection*
- 3) *Pre-Processing (Data Cleaning)*
- 4) *Transformation Data*
- 5) *Data Mining*
- 6) *Evaluation*

#### 3.1 Remove Data Tagihan

Data WP yang telah ada dipisahkan menggunakan *tools Pentaho*. *Pentaho* adalah kumpulan aplikasi *business intelligence (BI)* dan *data warehouse free open source software* yang berjalan diatas *platform java* (Putra, 2015).

Berdasarkan perwali (peraturan walikota) Palembang yang baru bahwa tagihan pajak bumi dan bangunan (PBB) dibawah Rp. 100.000 akan dibebaskan pembayarannya, maka dari itu tagihan yang dibawah Rp. 100.000 akan dihapuskan dari sumber data yang ada menggunakan *software Pentaho*. Setelah

dihilangkan setelah dihilangkan menggunakan *pentaho* menjadi sebanyak 1.647 data wajib pajak.

### 3.2 Data Selection

*Classification* adalah proses penemuan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya belum diketahui. (Vulandari, 2017)

Pada tahap ini *field* dari sumber data akan dipisahkan antara field yang tidak digunakan dan yang akan digunakan untuk penggalian data pada tahap *pre-processing* selanjutnya untuk mendapatkan *knowledge* yang diinginkan dengan menggunakan *tools rapidminer*.

Row No.	Class	NOP	Keluaran OP	Total NJOP	Total NJOP	Tgl Jatuh T.	Tanggal Res.
1631	Tidak Wajib	1671933010	5637735A	42200000	41600000	30-09-2017	14-02-2018 L.
1632	Tidak Wajib	1671933010	5637735A	90440000	47600000	30-09-2017	17-07-2017 L.
1633	Sangat Wajib	1671933004	14 LULU	20440000	38075000	30-09-2017	13-04-2017 L.
1634	Sangat Wajib	1671933002	12 LULU	155040000	160300000	30-09-2017	14-03-2017 L.
1635	Tidak Wajib	1671933010	5637735A	44500000	58300000	30-09-2017	04-08-2017 L.
1636	Tidak Wajib	1671933004	14 LULU	9840000	17900000	30-09-2017	16-08-2017 L.
1637	Sangat Wajib	1671933010	5637735A	72000000	30030000	30-09-2017	03-04-2017 B.
1638	Sangat Wajib	1671933004	14 LULU	146432000	19030000	30-09-2017	08-05-2017 B.
1639	Tidak Wajib	1671933004	14 LULU	26800000	99800000	30-09-2017	20-09-2017 L.
1640	Tidak Wajib	1671933004	14 LULU	58240000	42400000	30-09-2017	25-10-2017 L.
1641	Sangat Wajib	1671933010	5637735A	62400000	40980000	30-09-2017	26-04-2017 L.
1642	Tidak Wajib	1671933004	14 LULU	70400000	71400000	30-09-2017	21-07-2017 B.
1643	Sangat Wajib	1671933004	14 LULU	51200000	42400000	30-09-2017	06-04-2017 L.
1644	Sangat Wajib	1671933004	14 LULU	170100000	18900000	30-09-2017	12-05-2017 L.
1645	Sangat Wajib	1671933010	5637735A	18030000	28900000	30-09-2017	27-04-2017 B.
1646	Tidak Wajib	1671933004	14 LULU	46718000	60840000	30-09-2017	23-10-2017 B.
1647	Tidak Wajib	1671933004	14 LULU	224120000	100000000	30-09-2017	07-08-2017 L.

Gambar 2. Hasil Seleksi Data

*Selection* berfungsi untuk memilih data yang sesuai kebutuhan untuk mendapatkan pengetahuan, pada penelitian ini data yang sesuai untuk mendapatkan hasil yang diinginkan dan sesuai dengan prosesnya adalah NOP, kelurahan OP, total NJOP bumi, total NJOP bangunan, tanggal jatuh tempo, dan tanggal bayar serta ada tambahan field data yaitu estimasi dan class

sebagai akurasi untuk prediksi potensi kepatuhan.

### 3.3 Pre-Processing(Data Cleaning)

Data yang telah diseleksi pada tahapan sebelumnya masih memiliki data yang kosong, ganda dan *missing*. Selain itu juga, data yang tidak relevan itu juga lebih baik dibuang karena keberadaannya bisa mengurangi akurasi dari hasil data mining nantinya. Dengan proses data yang telahdiseleksi sebelumnya akan dilihat apakah pada data tersebut terdapat data yang kosong dan terdapat nilai *missing*, jika tidak terdapat data kosong dan *missing* data tersebut akan langsung diproses pada tahap tranformasi data, sedangkan jika terdapat data yang kosong dan *missing* data tersebut akan dibersihkan terlebih dahulu dengan menggunakan sebelum masuk pada tahap transformasi data.

Gambar 3. Hasil Pembersihan Data Kosong dan Missing

Gambar 3 merupakan hasil dari pembersihan data terhadap data kosong dan *missing* ternyata pada data PBB SU 2 tidak terdapat data yang *missing* dan kosong maka data wajib pajak tetap sebanyak 1.647 data.

### 3.4 Transformation Data

Pada tahapan ini data yang telah dipilih dan dibersihkan selanjutnya akan ditransformasikan kedalam bentuk-bentuk yang cocok dan sesuai dengan algoritma yang digunakan untuk *procedure* penggalian data dengan cara melakukan normalisasi data yang akan digunakan *operator normalize* untuk mengubah data awal menjadi bentuk yang cocok untuk dilakukan penggalian data menggunakan *algoritma naive bayes*, tahap ini perlu dilakukan untuk memperbaiki nilai akurasi yang didapat pada tahap akhir nantinya, jika tidak dilakukan *normalize* data maka hasil akurasi yang akan didapat akan menjadi kecil.

No	Class	NOP	Total NJOP	Kabupaten OP	Tgl Jatuh	Tanggal Bayar
1	Sangat Tepat	0	-0.204	13 ULU	30-09-2017	30-03-2017.1
2	Sangat Tepat	0	-0.202	16 ULU	30-09-2017	07-06-2017.1
3	Sangat Tepat	0	-0.211	16 ULU	30-09-2017	22-09-2017.1
4	Tepat Waktu	0	-0.180	16 ULU	30-09-2017	28-09-2017.1
5	Tepat Waktu	0	-0.177	16 ULU	30-09-2017	18-09-2017.3
6	Sangat Tepat	0	0.004	TANGGA EK.	30-09-2017	19-09-2017.1
7	Tepat Waktu	0	0.017	16 ULU	30-09-2017	28-09-2017.1
8	Sangat Tepat	0	-0.185	16 ULU	30-09-2017	08-04-2017.3
9	Tepat Waktu	0	-0.283	13 ULU	30-09-2017	25-09-2017.1
10	Tepat Waktu	0	-0.091	16 ULU	30-09-2017	15-09-2017.3
11	Tepat Waktu	0	-0.216	16 ULU	30-09-2017	13-09-2017.9
12	Sangat Tepat	0	-0.202	16 ULU	30-09-2017	05-06-2017.1
13	Tepat Waktu	0	-0.220	16 ULU	30-09-2017	24-07-2017.1
14	Tepat Waktu	0	-0.176	TANGGA EK.	30-09-2017	16-09-2017.1
15	Sangat Tepat	0	-0.164	16 ULU	30-09-2017	06-04-2018.1
16	Tepat Waktu	0	-0.229	16 ULU	30-09-2017	24-07-2017.1
17	Tepat Waktu	0	1.372	TANGGA EK.	30-09-2017	27-09-2017.1

Gambar 4. Hasil transformasi data

Pada gambar 4 dapat dilihat data yang telah dilakukan transformasi menjadi bentuk yang sesuai untuk dilakukan penggalian data pada tahapan selanjutnya, terdapat perubahan pada *record* dari *field* NOP, Total NJOP Bumi, dan Total NJOP Bangunan diubah menjadi bentuk yang cocok untuk dimining.

### 3.5 Data Mining

Pada tahap ini dilakukan 2 pengujian yaitu dengan melakukan perhitungan manual dan perhitungan dari *rapidminer* untuk mendapatkan hasil yang akurat.

#### 1) Perhitungan Manual Algoritma *naive bayes*

Berikut perhitungan manual dari algoritma *Naive Bayes* dengan menggunakan dataset training sebanyak 200 *record* dengan diambil acak dari data PBB kecamatan seberang ulu 2 yang telah dilakukan pembuangan data dengan tagihan kurang dari Rp. 100.000. Perhitungan manual algoritma *naive bayes* dilakukan menggunakan rumus sebagai berikut:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \dots \dots \dots (1)$$

Keterangan:

- P (H|X) = Probabilitas
- H = Hipotesis
- X = Bukti .

Sebelum dilakukan perhitungan manual harus ditentukan dulu estimasi dari setiap tanggal bayar dan tanggal jatuh tempo dengan melakukan perhitungan dan membuat peraturan untuk mendapatkan estimasi, maka aturan atau *rule* yang terbentuk adalah sebagai berikut:

- a. *If* tanggal bayar = januari 2017 *Then* Estimasi = 4
- b. *If* tanggal bayar = february 2017 *Then* Estimasi = 5
- c. *If* tanggal bayar = maret 2017 *Then* Estimasi = 6
- d. *If* tanggal bayar = april 2017 *Then* Estimasi = 7
- e. *If* tanggal bayar = mey 2017 *Then* Estimasi = 8
- f. *If* tanggal bayar = juni 2017 *Then* Estimasi = 9

- g. *If* tanggal bayar = juli 2017 *Then* Estimasi = 10
- h. *If* tanggal bayar = agustus 2017 *Then* Estimasi = 11
- i. *If* tanggal bayar = september 2017 *Then* Estimasi = 12
- j. *If* tanggal bayar = oktober 2017 *Then* Estimasi = 13
- k. *If* tanggal bayar = november 2017 *Then* Estimasi = 14
- l. *If* tanggal bayar = desember 2017 *Then* Estimasi = 15
- m. *If* tanggal bayar = januari 2018 *Then* Estimasi = 16
- n. *If* tanggal bayar = februari 2018 *Then* Estimasi = 17
- o. *If* tanggal bayar = maret 2018 *Then* Estimasi = 18
- p. *If* tanggal bayar = april 2018 *Then* Estimasi = 19
- q. *If* tanggal bayar = mey 2018 *Then* Estimasi = 20
- r. *If* tanggal bayar = juni 2018 *Then* Estimasi = 21
- s. *If* estimasi  $\leq 9$  *Then* estimasi(class) = Sangat Tepat Waktu
- t. *If* estimasi = 10 – 12 *Then* estimasi(class) = Tepat Waktu
- u. *If* estimasi  $> 12$  *Then* estimasi(class) = Tidak Tepat Waktu

Adapun dataset untuk melakukan perhitungan manual terdiri dari 200 *record* yang diambil acak dari data PBB kecamatan seberang ulu 2 yang telah dihapuskan data yang dibawah Rp. 100.000, dengan contoh data seperti pada tabel 1.

**Tabel 1. Dataset**

NOP	KELURAHAN OP	TGL JATUH TEMPO	TGL BAYAR	ESTIMASI
167103000500430030	16 ULU	30-09-2017	05-04-2017 09:43:08	SangatTepatWaktu
167103000300200570	13 ULU	30-09-2017	30-03-2017 12:17:31	SangatTepatWaktu
167103000300201200	13 ULU	30-09-2017	13-03-2017 09:13:13	SangatTepatWaktu
167103001000202380	SENTOSA	30-09-2017	31-03-2017 11:07:01	SangatTepatWaktu
167103001000203230	SENTOSA	30-09-2017	22-05-2017 08:50:58	SangatTepatWaktu
167103000501311450	16 ULU	30-09-2017	04-10-2017 12:14:56	TidakTepatWaktu
167103000400401220	14 ULU	30-09-2017	10-05-2017 11:26:25	SangatTepatWaktu
167103000500902900	16 ULU	30-09-2017	07-06-2017 11:08:53	SangatTepatWaktu
167103000500201740	16 ULU	30-09-2017	22-06-2017 10:59:30	SangatTepatWaktu
167103000500900070	16 ULU	30-09-2017	26-09-2017 10:27:49	TepatWaktu
167103000500203440	16 ULU	30-09-2017	14-08-2017 09:51:20	TepatWaktu
167103001000102170	SENTOSA	30-09-2017	03-04-2018 08:05:57	SangatTepatWaktu
167103000500201380	16 ULU	30-09-2017	12-05-2017 10:21:59	SangatTepatWaktu
167103000900601290	TANGGA TAKAT	30-09-2017	05-06-2017 12:59:23	SangatTepatWaktu
167103000901003540	TANGGA TAKAT	30-09-2017	19-05-2017 10:54:34	SangatTepatWaktu
167103000900402780	TANGGA TAKAT	30-09-2017	Belum Bayar	TidakTepatWaktu
167103000501107680	16 ULU	30-09-2017	Belum Bayar	TidakTepatWaktu
167103000501306870	16 ULU	30-09-2017	Belum Bayar	TidakTepatWaktu
167103000500500920	16 ULU	30-09-2017	Belum Bayar	TidakTepatWaktu
167103000501101320	16 ULU	30-09-2017	Belum Bayar	TidakTepatWaktu

Menghitung jumlah kelas dari estimasi berdasarkan klasifikasi yang terbentuk (*prior probability*), dimana jumlah tiap kelas masing-masing estimasi yang ada dibagi dengan jumlah seluruh data. C1 untuk kelas estimasi sangat tepat waktu, C2 untuk kelas tepat waktu dan C3 untuk kelas tidak tepat waktu:

- a. C1(Class Estimasi = ‘Sangat Tepat Waktu’) = jumlah “sangat tepat waktu” pada kolom Estimasi =  $97/200 = 0,48$
- b. C2 (Class Estimasi = “Tepat Waktu”) = jumlah “tidak tepat waktu” pada kolom Estimasi =  $48/200 = 0,24$
- c. C3 (Class Estimasi = “Tidak Tepat Waktu”) = jumlah “tidak tepat waktu” pada kolom Estimasi =  $55/200 = 0,27$

Kemudian Menghitung jumlah kasus yang sama pada setiap atribut dari kelas Estimasi (sangat tepat waktu , tepat waktu/tidak tepat waktu) berdasarkan masing-masing kelurahan OP (13 ULU, 14 ULU, 16 ULU TANGGA TAKAT dan SENTOSA) kemudian dibagi dengan seluruh jumlah atribut kelurahan OP masing-masing .

- a.  $P(\text{Kelurahan} = \text{"11 ULU"} \mid \text{Class Estimasi} = \text{"sangat tepat waktu"}) = 7/8 = 0,87$
- b.  $P(\text{Kelurahan} = \text{"11 ULU"} \mid \text{Class Estimasi} = \text{"tepat waktu"}) = 1/8 = 0,12$
- c.  $P(\text{Kelurahan} = \text{"11 ULU"} \mid \text{Class Estimasi} = \text{"tidak tepat waktu"}) = 0/8 = 0$
- d.  $P(\text{Kelurahan} = \text{"12 ULU"} \mid \text{Class Estimasi} = \text{"sangat tepat waktu"}) = 7/12 = 0,58$
- e.  $P(\text{Kelurahan} = \text{"12 ULU"} \mid \text{Class Estimasi} = \text{"tepat waktu"}) = 3/12 = 0,25$
- f.  $P(\text{Kelurahan} = \text{"12 ULU"} \mid \text{Class Estimasi} = \text{"tidak tepat waktu"}) = 2/12 = 0,16$
- g.  $P(\text{Kelurahan} = \text{"13 ULU"} \mid \text{Class Estimasi} = \text{"sangat tepat waktu"}) = 18/28 = 0,64$
- h.  $P(\text{Kelurahan} = \text{"13 ULU"} \mid \text{Class Estimasi} = \text{"tepat waktu"}) = 6/28 = 0,21$
- i.  $P(\text{Kelurahan} = \text{"13 ULU"} \mid \text{Class Estimasi} = \text{"tidak tepat waktu"}) = 4/28 = 0,14$
- j.  $P(\text{Kelurahan} = \text{"14 ULU"} \mid \text{Class Estimasi} = \text{"sangat tepat waktu"}) = 6/15 = 0,4$
- k.  $P(\text{Kelurahan} = \text{"14 ULU"} \mid \text{Class Estimasi} = \text{"tepat waktu"}) = 6/15 = 0,4$
- l.  $P(\text{Kelurahan} = \text{"14 ULU"} \mid \text{Class Estimasi} = \text{"tidak tepat waktu"}) = 3/15 = 0,2$
- m.  $P(\text{Kelurahan} = \text{"16 ULU"} \mid \text{Class Estimasi} = \text{"sangat tepat waktu"}) = 26/70 = 0,37$
- n.  $P(\text{Kelurahan} = \text{"16 ULU"} \mid \text{Class Estimasi} = \text{"tepat waktu"}) = 16/70 = 0,22$
- o.  $P(\text{Kelurahan} = \text{"16 ULU"} \mid \text{Class Estimasi} = \text{"tidak tepat waktu"}) = 28/70 = 0,4$
- p.  $P(\text{Kelurahan} = \text{"Tangga Takat"} \mid \text{Class Estimasi} = \text{"sangat tepat waktu"}) = 21/37 = 0,56$
- q.  $P(\text{Kelurahan} = \text{"Tangga Takat"} \mid \text{Class Estimasi} = \text{"tepat waktu"}) = 10/37 = 0,27$

- r.  $P(\text{Kelurahan} = \text{"Tangga Takat"} \mid \text{Class Estimasi} = \text{"tidak tepat waktu"}) = 6/37 = 0,16$
- s.  $P(\text{Kelurahan} = \text{"Sentosa"} \mid \text{Class Estimasi} = \text{"sangat tepat waktu"}) = 12/30 = 0,4$
- t.  $P(\text{Kelurahan} = \text{"Sentosa"} \mid \text{Class Estimasi} = \text{"tepat waktu"}) = 6/30 = 0,2$
- u.  $P(\text{Kelurahan} = \text{"Sentosa"} \mid \text{Class Estimasi} = \text{"tidak tepat waktu"}) = 12/30 = 0,4$

Selanjutnya kalikan Semua Hasil *Variable* yang didapat dari hasil  $P(\text{KELURAHAN OP} \mid \text{Class Estimasi})$  dengan sesama kelas estimasinya (C1, C2 dan C3), adalah sebagai berikut :

- a. Untuk semua atribut *Class Estimasi* = "sangat tepat waktu"  
 $P(X \mid \text{Class Estimasi} = \text{"sangat tepat waktu"}) = 0,87 \times 0,58 \times 0,64 \times 0,4 \times 0,37 \times 0,56 \times 0,4 = 0,01$
- b. Untuk semua atribut *Class Estimasi* = "tepat waktu"  
 $P(X \mid \text{Class Estimasi} = \text{"tepat waktu"}) = 0,12 \times 0,25 \times 0,21 \times 0,4 \times 0,22 \times 0,27 \times 0,2 = 0,04$
- c. Untuk semua atribut *Class Estimasi* = "tidak tepat waktu"  
 $P(X \mid \text{Class Estimasi} = \text{"tidak tepat waktu"}) = 0 \times 0,16 \times 0,14 \times 0,2 \times 0,4 \times 0,16 \times 0,4 = 0$   
 Lalu kalikan hasil yang didapat dari P (NOP | kelas estimasi) dengan hasil yang didapat dari hasil perkalian semua *variable* (C1, C2 dan C3) diatas :
  - a. Perkalian *priot probability* dengan semua atribut *Class Estimasi* = "sangat tepat waktu"  
 $P(C_i) \mid \text{Class Estimasi} = \text{"sangat tepat waktu"} \times P(X \mid \text{Class Estimasi} = \text{"sangat tepat waktu"})$



$$= 0,48 \times 0,01$$

$$= 0,0048$$

- b. Perkalian *prior probability* dengan semua atribut *Class Estimasi* = “tepat waktu”

$$P(C_i) | \text{Class Estimasi} = \text{“tepat waktu”} \times$$

$$P(X|\text{Class Estimasi} = \text{“tepat waktu”})$$

$$= 0,24 \times 0,04$$

$$= 0,0096$$

- c. Perkalian *prior probability* dengan semua atribut *Class Estimasi* = “tidak tepat waktu”

$$P(C_i) | \text{Class Estimasi} = \text{“tidak tepat waktu”} \times$$

$$P(X|\text{Class Estimasi} = \text{“tidak tepat waktu”})$$

$$= 0,27 \times 0 = 0$$

Dari hasil perhitungan  $P(X|C_i) \cdot P(C_i)$  didapat perbandingan dimana  $P(C_i | \text{Class Estimasi} = \text{“Tepat Waktu”}) \cdot P(X|\text{Class Estimasi} = \text{“tepat waktu”}) > P(C_i | \text{Class Estimasi} = \text{“Sangat Tepat Waktu”}) \cdot P(X|\text{Class Estimasi} = \text{“sangat tepat waktu”}) \& P(C_i | \text{Class Estimasi} = \text{“tidak tepat waktu”}) \cdot P(X|\text{Class Estimasi} = \text{“tidak tepat waktu”})$

Kesimpulan :

Perhitungan antara perkalian *Class Estimasi* “sangat tepat waktu”, *Class Estimasi* “tepat waktu”, dan *Class Estimasi* “tidak tepat waktu” menunjukkan bahwa nilai *Class Estimasi* = “tepat waktu” lebih besar dibandingkan *class Estimasi* “sangat tepat waktu” dan *class estimasi* “tidak tepat waktu”

## 2) Penerapan Algoritma *Naive Bayes* Pada *RapidMiner*

Tahapan implementasi algoritma *naive bayes* menggunakan *tools rapidminer* untuk menentukan potensi kepatuhan wajib pajak. Gambar 5 berikut merupakan hasil dari potensi kepatuhan wajib pajak dengan ditambahkan nilai

*confidence* sebagai hasil prediksi dari *tools rapidminer*.

Row No.	NCP	Class	predictedClass	confidence/Sangat Tepat Waktu	confidence/tepat Waktu	confidence/Tidak Tepat Waktu	Total NCP	Total NCP
1	0	Tepat Waktu	Tepat Waktu	0.991	0.993	0.006	-0.100	-0.270
2	0	Tepat Waktu	Tepat Waktu	0.991	0.992	0.007	-0.077	-0.066
3	0	Sangat Tepat	Sangat Tepat Waktu	0.990	0.992	0.008	-0.165	-0.173
4	0	Tepat Waktu	Tepat Waktu	0.991	0.993	0.006	-0.270	-0.137
5	0	Sangat Tepat	Sangat Tepat Waktu	0.992	0.992	0.006	-0.262	-0.123
6	0	Sangat Tepat	Sangat Tepat Waktu	0.991	0.992	0.007	-0.154	0.038
7	0	Tepat Waktu	Tepat Waktu	0.990	1.000	0.000	1.372	-0.277
8	0	Sangat Tepat	Sangat Tepat Waktu	0.992	0.998	0.000	0.987	-0.905
9	0	Tepat Waktu	Tepat Waktu	0.992	0.991	0.007	-0.103	-0.238
10	0	Sangat Tepat	Sangat Tepat Waktu	0.990	0.992	0.008	-0.140	-0.063
11	0	Tidak Tepat	Tidak Tepat Waktu	0.990	0.999	1.000	0.006	-0.115
12	0	Tepat Waktu	Tepat Waktu	0.991	0.996	0.003	0.150	0.111
13	0	Sangat Tepat	Sangat Tepat Waktu	0.990	0.992	0.008	-0.167	-0.065
14	0	Sangat Tepat	Sangat Tepat Waktu	0.992	0.992	0.006	-0.263	-0.102
15	0	Tepat Waktu	Tepat Waktu	0.991	0.993	0.005	-0.258	-0.205
16	0	Tidak Tepat	Tidak Tepat Waktu	0.990	0.990	0.998	-0.008	-0.104

Gambar 5. Hasil Potensi

## 3.6 Evaluation

Pada tahap ini pola-pola hasil akhir dari proses data mining akan dipresentasikan dalam bentuk visual untuk membantu *user* dalam mengerti pola tersebut.

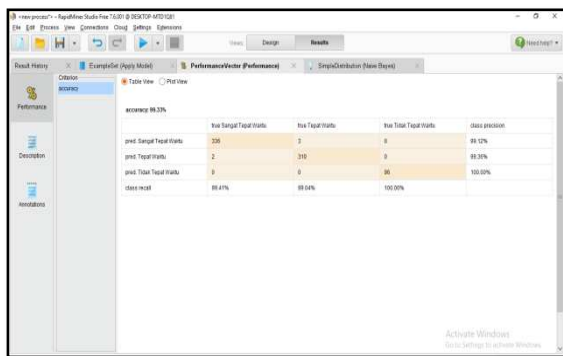
### 1) Akurasi Prediksi

Pada percobaan dengan algoritma *naive bayes* dengan menggunakan *tools rapidminer* diperoleh waktu eksekusi sebesar 1second dapat diartikan eksekusi berjalan cukup cepat. Hasil eksekusi dapat dilihat pada gambar 6berikut.

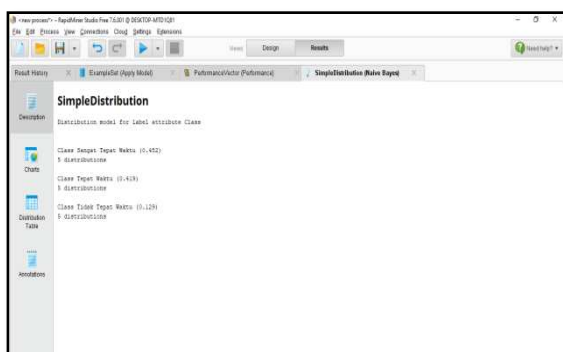
Process Name	Completed Date/Time	Execution Time
Process (1) results	Aug 31, 2018 8:30:16 AM	0 s
Process (1) results	Aug 31, 2018 8:34:13 AM	0 s
Process (1) results	Aug 31, 2018 8:35:37 AM	0 s
Process (1) results	Aug 31, 2018 8:37:16 AM	0 s
Process (1) results	Aug 31, 2018 8:41:36 AM	0 s
Process (2) results	Aug 31, 2018 8:51:14 AM	1 s
Process (2) results	Aug 31, 2018 8:52:35 AM	0 s

Gambar 6. Hasil eksekusi

Hasil akurasi model *naive bayes* menunjukkan bahwa nilai tingkat akurasi yang dihasilkan oleh algoritma *naive bayes* memiliki tingkat kekuatan yang sangat tinggi. Hal ini dibuktikan dengan hasil akurasi perhitungan mencapai 99.33%, nilai tersebut membuktikan bahwa model yang dibangun dapat digunakan untuk melakukan klasifikasi potensi kepatuhan wajib pajak, nilai yang tinggi tersebut juga disebabkan oleh kompleksnya data yang mengakibatkan model dapat memprediksi dengan akurat. Adapun hasil akurasi dapat dilihat pada gambar 7.



Gambar 7. Hasil Akurasi



Gambar 8 Model Distribusi

Dari gambar 8 dapat ditarik kesimpulan bahwa data PBB kecamatan seberang ulu 2 Palembang kelas sangat tepat waktu dengan nilai sebesar 0.382 lebih besar dari kelas tepat waktu dan kelas tidak tepat waktu. Adapun model

distribusi untuk label atribut kelas estimasi adalah sebagai berikut:

- Kelas sangat tepat waktu : 5 distribution
- Kelas tepat waktu : 5 distribution
- Kelas tidak tepat waktu : 5 distribution

## 2) Hasil pengujian

Berdasarkan penelitian yang telah dilakukan, telah didapatkan suatu informasi dan pengetahuan baru dari proses data mining dalam menentukan wajib pajak di kelurahan mana yang berpotensi tidak patuh atau tidak tepat waktu, dan kebanyakan wajib pajak mana yang patuh membayar iuran PBB yang didapat setelah dilakukan tahapan *knowledge discovery in database* (KDD) dari data PBB kecamatan seberang ulu 2 tahun 2017. Dari penelitian tersebut didapatkan hasil yang sesuai dengan tujuan untuk menghasilkan informasi yang baru tentang kepatuhan wajib pajak yang berpotensi tidak tepat waktu.

Dari proses perhitungan data mining menggunakan algoritma *naive bayes* dengan menerapkan teknik *classification* dan tingkat keakurasiannya, dihasilkan informasi berdasarkan perhitungan data PBB kecamatan seberang ulu 2 Palembang tahun 2017 menunjukkan kelas sangat tepat waktu dengan perkalian *prior probability* senilai 0,48, sedangkan kelas tidak tepat waktu dengan *prior probability* senilai 0,27 dan kelas tepat waktu dengan *prior probability* senilai 0,24 serta menurut perhitungan dari *tools rapidminer* kelurahan yang tidak tepat waktu ada di kelurahan 16 ulu dari sebagian wajib pajak dengan data sebanyak 0.437 dan berpotensi tidak tepat waktu terdapat pada kelurahan tangga takat

dengan data sebanyak 0.229 dengan akurasi sebesar 99.33% yang diakibatkan berpotensi tidak tepat waktu karena besarnya total NJOP bumi dan bangunan yang mengakibatkan tingginya jumlah tagihan menyebabkan sulitnya masyarakat untuk membayar iuran PBB pertahunnya

### 3) Hasil Analisis

Dari proses perhitungan data mining secara manual dan dengan implementasi pada *tools rapidminer* didapatkan hasil wajib pajak di setiap kelurahan yang ada pada kecamatan seberang ulu 2 Palembang patuh dalam membayar iuran PBB, didapatkan kelurahan yang berpotensi tidak tepat waktu terdapat pada sebagian wajib pajak kelurahan 16 ulu dengan data sebanyak 42 wajib pajak dan kelurahantangga takat dengan data sebanyak 22 wajib pajak. Dengan hasil penelitian ini diharapkan dapat menjadi bahan referensi instansi pemerintah dalam memberikan arahan terhadap kelurahan tersebut agar menjadi patuh dalam membayar iuran PBB.

Adapun data wajib pajak dengan estimasi kepatuhannya sangat tepat waktu dapat dilihat pada tabel 2. Jumlah pada tabel 2 ini didapat dari jumlah seluruh kelas estimasi sangat tepat waktu pada dataset yang digunakan.

**Tabel 2. Hasil Pengujian**

Estimasi	Kelurahan	Jumlah	Prior Probability
Sangat Tepat Waktu	16 ULU	0.405	0.361
	TANGGA	0.246	0.142
	TAKAT		
	SENTOSA	0.121	0.139
	14 ULU	0.109	0.156
	13 ULU	0.077	0.125
	12 ULU	0.024	0.028
	11 ULU	0.018	0.049

**Tabel 3. Hasil Tepat Waktu**

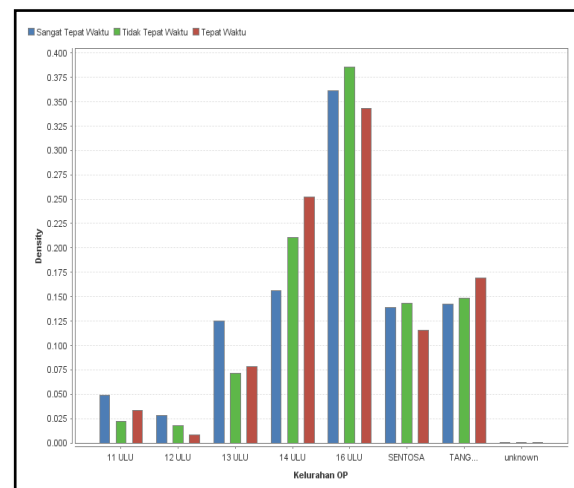
Estimasi	Kelurahan	Jumlah	Prior Probability
Tepat Waktu	16 ULU	0.434	0.386
	TANGGA	0.173	0.148
	TAKAT		
	14 ULU	0.150	0.211
	SENTOSA	0.141	0.143
	13 ULU	0.073	0.072
	11 ULU	0.022	0.022
	12 ULU	0.006	0.0018

**Tabel 4. Hasil Tidak Tepat Waktu**

Estimasi	Kelurahan	Jumlah	Prior Probability
Tidak Tepat Waktu	16 ULU	0.437	0.343
	TANGGA	0.229	0.169
	TAKAT		
	SENTOSA	0.146	0.116
	14 ULU	0.104	0.252
	13 ULU	0.073	0.079
	11 ULU	0.010	0.033
	12 ULU	0.000	0.008

### 4) Knowledge Representation

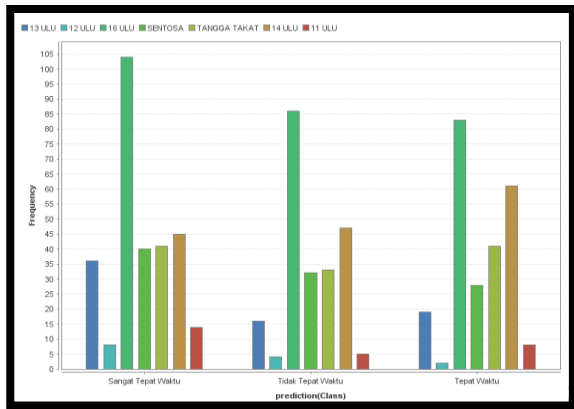
Ini merupakan tahapan akhir dimana pengetahuan yang telah ditemukan secara visual ditampilkan kepada user dengan berbagai visualisasi seperti *chart*, *histogram*, dan *pareto*. Adapun visualisasi tersebut dapat dilihat pada gambar 9, 10, dan 11.



**Gambar 9. Chart**

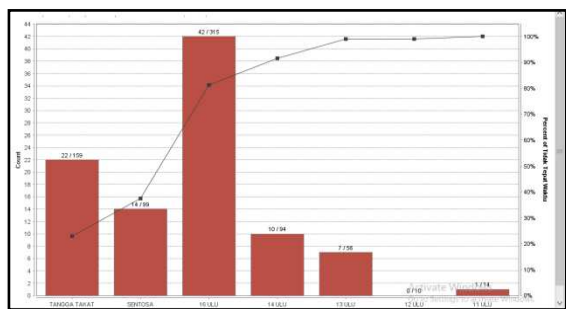
Gambar 9 memperlihatkan data wajib pajak kelurahan 16 ulu dengan presentase

diagram tertinggi yang memiliki data terbanyak diantara kelurahan lain dengan *class* estimasi tertinggi tidak tepat waktu mencapai >0.375.



**Gambar 10. Histogram**

Pada gambar 10 memperlihatkan *class* estimasi sangat tepat waktu, tepat waktu dan tidak teapat waktu tertinggi terdapat pada kelurahan 16 ulu dan kelurahan 14 ulu



**Gambar 11. Pareto**

Pada gambar 11 memperlihatkan *class* tidak tepat waktu tertinggi terdapat pada kelurahan 16 ulu dan tangga takat.

#### 4. SIMPULAN

Berdasarkan perhitungan data mining menggunakan algoritma *naive bayes* pada teknik *classification* dapat ditarik kesimpulan bahwa

kelas estimasi pembayaran wajib pajak sangat tepat waktu lebih besar dari pada kelas estimasi tepat waktu dan tidak tepat waktu.

Dari hasil *observasi* terhadap data PBB kecamatan seberang ulu 2 Palembang melalui teknik *classification* menggunakan algoritma *naive bayes* didapatkan kelurahan yang berpotensi tidak tepat waktu terdapat pada sebagian wajib pajak kelurahan 16 ulu dengan data 0.437 dan kelurahan tangga takat dengan data sebanyak 0.229 yang diakibatkan oleh besarnya jumlah total NJOP bumi dan bangunan yang membuat semakin tingginya jumlah tagihan menyebabkan sulitnya masyarakat membayar iuran PBB pertahunnya.

*Tools rapidminer* berhasil menentukan potensi kepatuhan wajib pajak per kelurahan mana yang berpotensi tidak tepat waktu pada tahun 2017 dengan akurasi sebesar 99.33% dari jumlah seluruh data wajib pajak PBB kecamatan seberang ulu 2 Palembang sebesar 1.467 wajib pajak setelah dilakukan penghapusan tagihan dibawah Rp. 100.000

#### DAFTAR RUJUKAN

Nofriansyah, Dicky. 2014. *Konsep Data Mining VS Sistem Pendukung Keputusan*. Deepublish. Yogyakarta.

Sugiyono. 2013. *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D*. Alfabeta. Bandung.

Vulandari, Retno Tri. 2017. *Data Mining Teori Dan Aplikasi RapidMiner*. Gova Media. Yogyakarta.

<http://www.pajak.go.id/sites/default/files/BookletPBB.pdf>.